Time-to-birth prediction models and the influence of expert opinions

Gilles Vandewiele[1],
Isabelle Dehaene[2], Olivier Janssens[1], Femke Ongenae[1], Femke De Backere[1],
Filip De Turck[1], Kristien Roelens[2], Sofie Van Hoecke[1], and Thomas Demeester[1]

[1] IDLab, Ghent University – imec
Technologiepark-Zwijnaarde 126, Ghent, Belgium
`{name}.{familyname}@ugent.be`
[2] Department of Gynecology and Obstetrics, Ghent University Hospital
Corneel Heymanslaan 10, Ghent, Belgium
`{name}.{familyname}@uzgent.be`

Abstract. Preterm birth is the leading cause of death among children under five years old. The pathophysiology and etiology of preterm labor are not yet fully understood. This causes a large number of unnecessary hospitalizations due to high–sensitivity clinical policies, which has a significant psychological and economic impact. In this study, we present a predictive model, based on a new dataset containing information of 1,243 admissions, that predicts whether a patient will give birth within a given time after admission. Such a model could provide support in the clinical decision-making process. Predictions for birth within 48 hours or 7 days after admission yield an Area Under the Curve of the Receiver Operating Characteristic (AUC) of 0.72 for both tasks. Furthermore, we show that by incorporating predictions made by experts at admission, which introduces a potential bias, the prediction effectiveness increases to an AUC score of 0.83 and 0.81 for these respective tasks.

Keywords: preterm birth·clinical decision support·eHealth

1 Introduction

Preterm birth, defined as birth before 37 weeks of gestational age, is the leading cause of death among children younger than five years, according to the World Health Organization (WHO) [9]. In Flanders, the average prevalence rate amounts to 7%. Furthermore, for tertiary care centers in Flanders this can be significantly higher, as for example 18% of the deliveries in Ghent University Hospital are preterm. Today, a patient at risk of preterm birth, is often hospitalized in order to take measures that ameliorate the short- and long-term outcome for the neonate. These measures include the administering of antenatal corticosteroids (ACS) for fetal lung maturation, often under tocolysis for labor arrest. Unfortunately, these measures can have short- and long-term maternal and offspring side effects and should therefore only be taken when imminent birth is expected [14]. Moreover, the societal and personal psychological and economical burden related to the hospitalization of these patients, should not be underestimated. For example, the costs associated with preterm birth in the USA in 2005 were estimated around \$51,600 per infant [3]. Currently, the pathophysiology and etiology of preterm

labor are not yet fully understood, making it hard for experts to accurately determine whether a patient will give birth at term or not. As is often the case in medical domains, the sensitivity of a policy is typically considered more important than its specificity, resulting in a high number of false positives, or unnecessary hospitalizations.
Predictive machine learning models have been applied to numerous medical use cases [22]. The prior research on predictive models for preterm birth risk is shown in Table 1. The most important difference w.r.t. the presented study is the incorporation of expert opinions within the model. We assess the added value of a model that predicts the time-to-birth, based on a simple, interpretable logistic regression model. The input to the model consists of structured clinical variables which are available shortly after the patient's admission to the hospital, e.g. the gestational age at intake, the patient's BMI and how long membranes have been ruptured, and indications given by domain experts at admission.

Study	Dataset	Target	Results
[18]	170 patients w/ preterm labor; 24w $\leq$ gest. $\leq$ 34w; singleton preg.	birth $\leq$ 37w	AUC: 0.81
[5, 19]	58,807 singleton pregnancies; 20w $\leq$ gest. $\leq$ 25w	4 categories	AUC: 0.65 (mild) – 0.92 (extreme)
[17]	3 million singleton, 105,000 twin pregnancies and 4,000 triplet pregnancies	birth $\leq$ 32w	AUC: 0.65 (> 1 fetus) – AUC 0.73 (singleton)
[2]	906 patients (2 datasets) w/ preterm labor; 22w $\leq$ gest. $\leq$ 32w	birth $\leq$ 32w & birth in 48h	AUC: 0.73 (within 48h), AUC 0.72 ($\leq$ 32w)
[20]	142 singleton pregn. w/ preterm labor & intact membranes; 22w $\leq$ gest. < 34w	birth within 7d	AUC: 0.88
[16]	1.5 million singleton pregnancies; 22w $\leq$ gest.	birth $\leq$ 37w	AUC: 0.63
[4]	33,370 singleton pregnancies w/ preterm labor $\leq$ 34w; data collected 11w $\leq$ gest. $\leq$ 13w	birth $\leq$ 34w	AUC: 0.67
[15]	31,834 singleton pregnancy; data collected 11 $\leq$ gest. $\leq$ 14w; 24w $\leq$ gest. $\leq$ 37w	birth $\leq$ 37w	18.4% sensitivity, 97.1% specificity
[1]	2,699 patients; data from both first and second trimester; birth after 20w	birth $\leq$ 37w	AUC: 0.70
[13]	191 singleton pregnancies; only biomarker data; 24w $\leq$ gest.	birth $\leq$ 36w	AUC: 0.66 – 0.89
[23]	3,073 singleton pregnancies; multiple admissions	birth $\leq$ 37w	45.4%–59.4% sensitivity, 71.9%–84% specificity at third admission
[10]	617 patients with preterm labor; 22w $\leq$ gest. $\leq$ 32w	birth in 48h & birth $\leq$ 32w	AUC: 0.8 (birth within 48h), 0.85 (birth $\leq$ 32w)
[21]	600 singleton pregnancies from 10 centers; 24w $\leq$ gest. $\leq$ 34w	birth within 7d	AUC: 0.95
[7]	166 women with preterm labor; 24w $\leq$ gest. $\leq$ 31w	birth within 7d and 14d	AUC: 0.63
[6]	3,012 symptomatic women; 24w $\leq$ gest. $\leq$ 28w or readmission before 35w	birth within 7d	AUC: 0.724
[24]	prospective; 355 women with preterm labor; 24w $\leq$ gest. $\leq$ 34w	birth within 7d	100% sensitivity, 92.3% specificity
[11]	2,540 women; 24w $\leq$ gest.; data collected gest. $\leq$ 16w	birth $\leq$ 37w & birth $\leq$ 34w	AUC: 0.54 – 0.67 (37w); AUC: 0.56 – 0.70 (34w)
This study	**1,243 high-risk admissions; 24w $\leq$ gest. $\leq$ 37w**	**birth within 48h and 7d**	**AUC: 0.83 (48h); AUC: 0.81 (7d)**

Table 1: Predictive modeling studies for tasks related to birth prediction.

2 Methodology

2.1 Data collection and filtering

The dataset used within this study consists of data collected from patients admitted to the Department of Gynecology and Obstetrics at Ghent University Hospital, between 2012 and 2017. In total, 3,611 women were admitted in that period, corresponding to 4,332 pregnancies and 5,030 admissions. From these, 1,243 pregnancy-related admissions, corresponding to 1,145 high-risk pregnancies of 1,056 women, occurring between 24 and 37 weeks of gestation, were used in the proposed pipeline. The reason for excluding other admissions is because the clinical use of our model is limited for these type of admissions. Patients at a gestational age less than 24 weeks are not included, since neonatal intensive care is not started before this term in Ghent University Hospital. Patients arriving at the hospital after 37 weeks of gestation are no longer at risk for preterm birth and thus do not require potential preventive measures.

2.2 Predictive variables

From the data we extract: number of fetuses, age (mother), gravidity, parity, length (mother), weight (mother), BMI, gestational age at admission, duration ruptured membranes, method of conception, smoking history, alcohol usage, drug usage, history of cesarean section, race (mother), and admission indications. This list of variables has been constructed in consultation with domain experts. The admission indications are keywords that can either be objective observations including 'blood loss' and 'stomach ache', or more subjective keywords of experts such as 'imminent partus prematurus'. The latter type of keywords can include indications of what the expert expects to happen, i.e. an expert opinion. Hence, such keywords introduce a potential bias which could cause the model to simply repeat the predictions of an expert. This however does not need to be the case, especially if it turns out that the prediction implied by such keywords does not always hold true. In fact, these subjective expert predictions are recorded directly after the patient's intake, and we therefore propose to investigate the predictions of a model in which these expert opinions are actually used as highly informative features.

2.3 Data processing and modeling

Before feeding the data to a machine learning model, all variables were first transformed to a numerical form. To achieve this, categorical variables were one-hot-encoded and a bag-of-terms was constructed for each patient based on her listed keywords. This bag-of-terms is a k-dimensional binary vector, with k being the number of available keywords in the training set, and each value indicating the presence of a certain keyword. In our study, k is equal to 30. Afterwards, the processed data was fed to logistic regression classifiers to solve two tasks, corresponding to threshold values chosen in consultation with experts, as they are the bounds between which the effect of corticosteroids is thought to be optimal [12]. On the one hand, we will predict whether birth will occur within 48 hours after admission (***Task 1***), while on the other hand we make the prediction for birth within 7 days (***Task 2***).

3 Results

To assess the predictive performance of the proposed model, we measured different metrics using five-fold cross-validation, based on the patient identifiers rather than individual admissions. First, we report the classifier's accuracy. As accuracy does not provide a good indication of the model's predictive performance in the scenario of imbalanced data, we also report the specificity, sensitivity and Diagnostic Odds Ratio (DOR) obtained from the confusion matrices of our classifiers, and the Area Under the Curve

(AUC) of the Receiver Operating Characteristic curve (ROC) score. Table 2 summarizes the predictive performances for the models with and without inclusion of the expert predictions. The table lists the mean ± std over the five folds. From these results, we can conclude that ***Task 1*** seems to be a slightly easier prediction task. We hypothesize that this is due to the fact that patients that would give birth within 48 hours often have more distinctive symptoms. Further, we see that incorporating the biased expert predictions in our model results in a considerable increase in predictive performance. It should be noted that the sensitivity values are rather low, since we did not optimize specifically for this and kept the default decision boundary, i.e., at a predicted probability level of 0.5.

Pred.	Task	Accuracy	Specificity	Sensitivity	DOR	AUC
Without	$\leq$ 48h	0.74±0.02	0.94±0.02	0.32±0.05	9.3±5.3	0.72±0.02
	$\leq$ 7d	0.67±0.02	0.83±0.05	0.48±0.08	5.1±1.5	0.72±0.04
With	$\leq$ 48h	0.80±0.02	0.90±0.01	0.60±0.05	14.5±4.3	0.83±0.03
	$\leq$ 7d	0.76±0.02	0.83±0.03	0.67±0.02	10.7±2.7	0.81±0.02

Table 2: The predictive performances of a model that predicts whether a patient will give birth (i) within 48 hours, or (ii) within 7 days. Model without and model with inclusion of admission indication keywords as input features.

4 Conclusion

A simple and interpretable logistic regression model was presented to assess the time-to-birth of a patient upon admission to the hospital. Preliminary results show the positive impact of incorporating expert opinions within the model. Future work includes applying survival analysis to directly predict the time-to-birth, as opposed to a dichotomized target.

Acknowledgements: This research is funded by imec, the PRETURN clinical trial (B670201836255, EC/2018/0609) and a PhD SB fellow scholarship of FWO (1S31417N).

References

1. Alleman, B.W., et al.: A proposed method to predict preterm birth using clinical data, standard maternal serum screening, and cholesterol. AJOG **208**(6), 472–e1 (2013)
2. Allouche, M., et al.: Development & validation of nomograms for predicting ptb. AJOG **204**(3), 242 (2011)
3. Behrman, R.E., Butler, A.S., et al.: Societal costs of preterm birth (2007)
4. Beta, J., et al.: Prediction of spontaneous preterm delivery from maternal factors, obstetric history and placental perfusion and function at 11–13 weeks. Prenatal diagnosis **31**(1), 75–83 (2011)
5. Celik, E., et al.: Cervical length and obstetric history predict sptb: development & validation of a model to provide individualized risk assessment. UOG **31**(5), 549–554 (2008)
6. De Silva, D.A., et al.: Timing of delivery in a high-risk obstetric population: a clinical prediction model. BMC pregnancy and childbirth **17**(1), 202 (2017)
7. García-Blanco, A., et al.: Can stress biomarkers predict preterm birth in women with threatened preterm labor? Psychoneuroendocrinology **83**, 19–24 (2017)

8. Kayem, G., et al.: Prediction of clinical infection in women with preterm labour with intact membranes: a score based on ultrasonographic, clinical and biological markers. Eur. J. Obstet. Gynecol. Reprod. Biol. **145**(1), 36–40 (2009)
9. Liu, L., et al.: Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the sustainable development goals. The Lancet **388**(10063), 3027–3035 (2016)
10. Mailath-Pokorny, M., et al.: Individualized assessment of preterm birth risk using two modified prediction models. Eur. J. Obstet. Gynecol. Reprod. Biol. **186**, 42–48 (2015)
11. Meertens, L.J., et al.: Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation. Acta obstetricia et gynecologica Scandinavica (2018)
12. Melamed, N., et al.: Association between antenatal corticosteroid administration-to-birth interval and outcomes of preterm neonates. Obstetrics & Gynecology **125**(6), 1377–1384 (2015)
13. Menon, R., et al.: Multivariate adaptive regression splines analysis to predict biomarkers of spontaneous preterm birth. Acta obstetricia et gynecologica Scandinavica **93**(4), 382–391 (2014)
14. Roelens, K., et al.: Prevention of preterm birth in women at risk: Selected topics. Tech. rep., Belgian Health Care Knowledge Centre (KCE) (2014)
15. Sananes, N., et al.: Prediction of spontaneous preterm delivery in the first trimester of pregnancy. Eur. J. Obstet. Gynecol. Reprod. Biol. **171**(1), 18–22 (2013)
16. Schaaf, J.M., et al.: Development of a prognostic model for predicting spontaneous singleton preterm birth. Eur. J. Obstet. Gynecol. Reprod. Biol. **164**(2), 150–155 (2012)
17. Tan, H., et al.: Early prediction of preterm birth for singleton, twin, and triplet pregnancies. Eur. J. Obstet. Gynecol. Reprod. Biol. **131**(2), 132–137 (2007)
18. Tekesin, I., et al.: Evaluation and validation of a new risk score (cleopatra score) to predict the probability of premature delivery for patients with threatened preterm labor. UOG **26**(7), 699–706 (2005)
19. To, M., et al.: Prediction of patient-specific risk of early preterm delivery using maternal history and sonographic measurement of cervical length. UOG **27**(4), 362–367 (2006)
20. Tsiartas, P., et al.: Prediction of spontaneous preterm delivery in women with threatened preterm labour: a prospective cohort study of multiple proteins in maternal serum. BJOG **119**(7), 866–873 (2012)
21. Van Baaren, G.J., et al.: Risk factors for preterm delivery: do they add to fetal fibronectin testing and cervical length measurement in the prediction of preterm delivery in symptomatic women? Eur. J. Obstet. Gynecol. Reprod. Biol. **192**, 79–85 (2015)
22. Vandewiele, G., et al.: A decision support system to follow up and diagnose primary headache patients using semantically enriched data. BMC medical informatics and decision making **18**(1), 98 (2018)
23. Vovsha, I., et al.: Predicting preterm birth is not elusive: Machine learning paves the way to individual wellness. In: 2014 AAAI Symposia (2014)
24. Watson, H., et al.: Quipp app: a safe alternative to a treat-all strategy for threatened preterm labor. UOG **50**(3), 342–346 (2017)